

Customer Segmentation Analysis Based on Demographics and Behavioral Attributes

Team Members

- Rithwik Gilla, 50559606
- Sri Naga Charan Gampala, 50559716
- Peddireddy Venkata Krishna Reddy, 50560191
- Kotesbabu Bommana, 50559679

Introduction (Motivation and Significance)

Customer segmentation is a crucial aspect of marketing and business strategy, allowing businesses to tailor their products and services to meet the diverse needs of their customer base. The main aim of this project is to track the consumer behavior in today's market driven by various factors.

Methods

We have used R, a statistical and a powerful programming language for this project. Some of the methods that we have used on the data to implement the segmentation are: **Data Preprocessing, Exploratory Data Analysis (EDA) and Clustering Algorithms**, where the data is cleaned and prepared the data for further analysis, utilizing visualizations to understand the distribution of variables and identify patterns within the data.

We have taken both the train and test datasets from Kaggle website which has various customer behavior metrics.

Results

K-means Clustering:

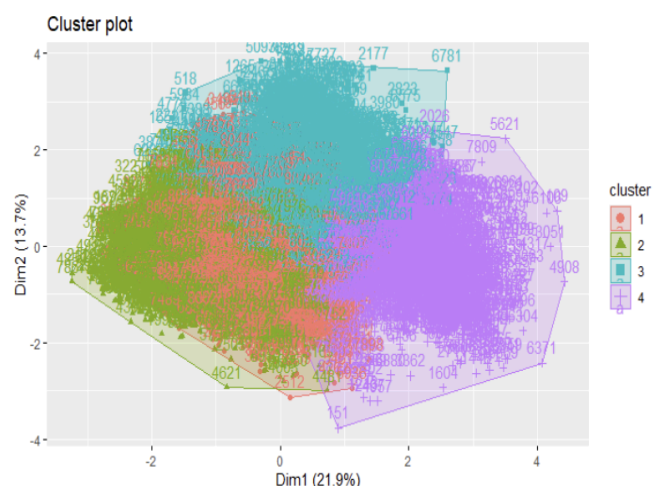
The K-means algorithm identified **4 clusters** with the following characteristics:

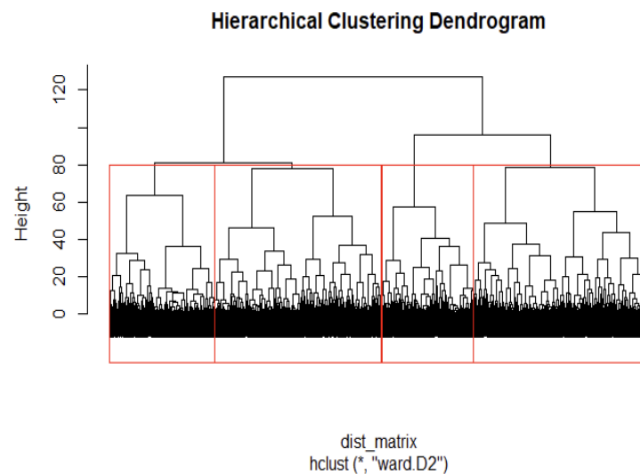
- Cluster 1:** 1790 members (low spending score),
- Cluster 2:** 2002 members (moderate spending score),
- Cluster 3:** 1834 members (high spending score),
- Cluster 4:** 1504 members (very high spending score)

Hierarchical Clustering:

The Hierarchical Clustering method also identified **4 clusters**:

Cluster 1: 1426 members, **Cluster 2:** 1833 members, **Cluster 3:** 1975 members, **Cluster 4:** 1896 members. The image below is the clustering dendrogram.



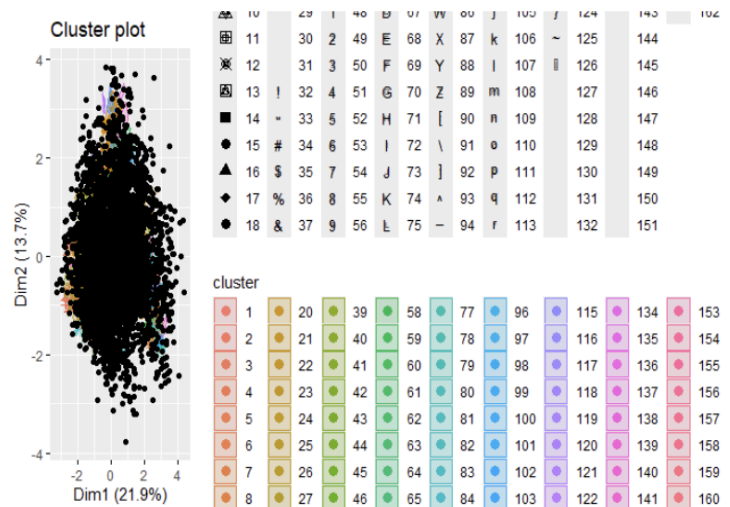


DBSCAN:

The DBSCAN algorithm identified **163 clusters** with **3,865 noise points**. This indicates that many data points did not fit well into any cluster, suggesting a high level of variability in the data.

Conclusion

Based on the results, **K-means clustering** appears to be the most effective method for this dataset, as it produced clear and actionable segments with distinct spending behaviors. Hierarchical clustering showed potential but lacked clear separation, while DBSCAN struggled with noise points. As consumer preferences continue to evolve, ongoing analysis and adaptation will be essential for maintaining a competitive edge in the market.



References:

- International Journal of Science and Research Archive. (2023). *Customer segmentation analysis for improving sales using clustering*. Retrieved from Semantic Scholar.
- ResearchGate. (n.d.). *Customer mobile behavioral segmentation and analysis in telecom using machine learning*. (Summarizing studies conducted between 2015 and 2019).
- IEEE Xplore. (n.d.). *Segmentation of online customers based on demographics and buying behavior*. (Publication year unavailable).
- SpringerLink. (n.d.). *Market segmentation and targeted marketing using machine learning models*. (Publication year unavailable).